International Journal of Computer Science and Innovation
Vol. 2015, no. 2, pp. 97-101
ISSN: 2458-6528

Infinity Sciences

Position Weight Matrix in Data Transfer

M. Yamuna

SAS, VIT University, Vellore
Email: *myamuna@vit.ac.in*

Abstract- *In communication systems, many encryption methods are developed, yet safe transfer of data is always a question and new methods are always proposed. A position weight matrix is very rarely used for data transfer. In this paper a method of data safe transfer using position weight matrix is proposed.*

Keywords**:** *Encryption; Decryption; Position weight matrix.*

I. INTRODUCTION

As communication methods change and improve day by day, the probability of them being hacked also increases. So safe transfer of data is always a standing problem. Various methods are proposed to improve the safety of the transmitted details. Various techniques like Steagnography, image encryption. The most common methods are symmetric, asymmetric and hashing method [1]. Better the difficulty of decrypting better the methods are. In [2] a review on encryption using different techniques is provided. In [3] a review on current algorithms used in encryption is discussed. In [4] a literature review on cryptography and network security is provided.

Matrices are important in encryption for it is computational friendly. Moreover it can be of any size and hence can be used to store large data. Various encryption techniques are developed based on matrices. In [5] a new method for encode and decode the data by using shared secret key, session key and intermediate key and matrix representations. In [6] a way of using positions of text represented in matrices for cryptography is introduced. The encryption system used a matrix to store the text entered by the sender in the form of their positions, using an algorithm to encrypt these values. In [7] polybit shuffling encryption and decryption algorithm based on N – dimensional encryption decryption matrix which is an attempt to improve over the classical playfair cipher is discussed. In [8], a novel approach for encryption of test messages suing playfair cipher 6y 6 matrix with four iteration steps is provided.

Matrices play an important role and hence new methods are continuously determined. Position weight matrix is in general used to represent details of DNA sequences and very rarely used in data transfer. In this paper a method of encrypting a message as a position weight matrix is proposed. This method does not use the routine matrix properties like XOR operations, bit operations; neither does it use the traditional methods of encryption, but uses biological data for encryption and decryption. This creates an impression that the matrix carries biological data rather than an encrypted message.

Position Weight Matrix

A position weight matrix has one row for each symbol of the alphabet A T G C in the DNA sequence. 4 rows for nucleotides in DNA sequences or 20 rows for amino acids in protein sequences. It also has one column for each position in the pattern. For example, given the following DNA sequences [9]

GAGGTAAAC
TCCGTAAGT
CAGGTTGGA
ACAGTCAGT
TAGGTCATT
TAGGTACTG
ATGGTAACT
CAGGTATAC
TGTGTGAGT
AAGGTAAGT

The corresponding PFM

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix} \begin{bmatrix} 3 & 6 & 1 & 0 & 0 & 6 & 7 & 2 & 1 \\ 2 & 2 & 1 & 0 & 0 & 2 & 1 & 1 & 2 \\ 1 & 1 & 7 & 10 & 0 & 1 & 1 & 5 & 1 \\ 4 & 1 & 1 & 0 & 10 & 1 & 1 & 2 & 6 \end{bmatrix}.$$

II. METHODS

A. Proposed Encryption Scheme

A position weight matrix has a special property that the sum of each column is same. The row labels are A, C, G, T. Each row represents these nucleotides. We now use this position of the nucleotides by assigning same kind of labels to the columns. In the resulting matrix, the possible entries are as seen in the following matrix.

	A	T	G	C
A	AA	AT	AG	AC
C	CA	CT	CG	CC
G	GA	GT	GG	GC
T	TA	TT	TG	TC

We use the entries with same nucleotides for our encryption purpose, ie., we use AA, GG, CC, TT for our purpose. We then use random values to create a resulting matrix so that each column entry is the same, so that it looks like a position weight matrix. For normal encryption we use the 26 alphabets and blank space. We use the following chart for encoding the message in the first stage.

Blank Space	**A**	**B**	**C**	• • •	**X**	**Y**	**Z**
↕	↕	↕	↕		↕	↕	↕
0	**1**	**2**	**3**		**24**	**25**	**26**

Encoding Chart

B. Algorithm of the Encryption Decryption Matrix

Let the message S: m1 m2… mk to be encrypted be of length k.

Step 1 Convert each character in the message into a sequence S1: n1 n2… nk of numbers of length k. Let the largest value of these k numbers be L.

Step 2 Construction of Encryption Matrix

- We choose the number of columns in the message matrix M to be k. For our example the number of columns is 9.
- Choose the number of rows as 4.
- Assign row and column labels as follows
  Row label: Assign the row labels as A C G T.

  Column Label: Assign the column labels as A T G C A T G C repeatedly (Note that A T G C is repeated again and again).

- Assign the value of n1 at position AA, n2 at position TT, n3, at position GG, n4 at position CC, n5 at position AA until all the k values are exhausted.
- Assign random values to the remaining entries in the matrix so that the sum of each column is L to generate the message matrix M.

Step 3 Send the matrix M to the receiver.

C. Illustration of the Algorithm

For example let S: **GOOD LUCK** be the message to be encrypted. Then length of S = k = 9. For our example the sequence obtained is from the encryption chart is

S1: 7 15 15 4 0 12 21 3 11. The largest value L = 21.

The column labels are A T G C A T G C A. The matrix of labels only for our example is seen in the following matrix.

	A	T	G	C	A	T	G	C	A
A	AA	AT	AG	AC	AA	AT	AG	AC	AA
C	CA	CT	CG	CC	CA	CT	CG	CC	CA
G	GA	GT	GG	GC	GA	GT	GG	GC	GA
T	TA	TT	TG	TC	TA	TT	TG	TC	TA

String S1 is 7 15 15 4 0 12 21 3 11. Assign value 7 to position AA of column 1, 15 to position TT of column 2,…, 11 to position AA of column to generate the following matrix.

	A	T	G	C	A	T	G	C	A
A	7	AT	AG	AC	0	AT	AG	AC	11
C	CA	CT	CG	2	CA	CT	CG	3	CA
G	GA	GT	15	GC	GA	GT	21	GC	GA
T	TA	15	TG	TC	TA	12	TG	TC	TA

Assign random values to the remaining entries in the matrix so that the sum of each column is L = 21 to generate the message matrix

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix}\begin{bmatrix} 7 & 3 & 3 & 9 & 0 & 3 & 0 & 9 & 11 \\ 7 & 3 & 2 & 2 & 7 & 4 & 0 & 3 & 4 \\ 5 & 0 & 15 & 6 & 7 & 2 & 21 & 0 & 4 \\ 2 & 15 & 1 & 4 & 7 & 12 & 0 & 9 & 2 \end{bmatrix}$$

Send the matrix M to the receiver.

Note that the final matrix looks like a position weight matrix.

Decryption is done by reversing the procedure.

Suppose the received message is

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix}\begin{bmatrix} 20 & 6 & 9 & 2 & 11 & 0 \\ 0 & 3 & 4 & 14 & 3 & 1 \\ 0 & 3 & 1 & 2 & 2 & 0 \\ 0 & 8 & 6 & 2 & 4 & 19 \end{bmatrix}$$

Assigning row and column labels the resulting matrix is

$$M = \begin{matrix} & A & T & G & C & A & T \\ A \\ C \\ G \\ T \end{matrix}\begin{bmatrix} 20 & 6 & 9 & 2 & 11 & 0 \\ 0 & 3 & 4 & 14 & 3 & 1 \\ 0 & 3 & 1 & 2 & 2 & 0 \\ 0 & 8 & 6 & 2 & 4 & 19 \end{bmatrix}$$

Picking the values of AA, TT, GG, CC, AA, TT from columns 1 to 6 the sequence S1 generated is 20 8 1 14 11 19. From the encryption chart the message is decoded as **THANKS**.

D. Discussion

A position weight matrix is a matrix which carries information about DNA sequences and in not used in general for encryption techniques. This is the main advantage of the method, since it makes decryption difficult since the routine methods are not used. It used simple method that the entries with same nucleotides are used for carrying information regarding the encrypted message. The remaining entries are fake. Also the entries with same nucleotides do not occur in the same positions and rather they occur in different positions. Moreover the column and row labels are user defined. Hence it can vary with different users. Also we have repeated the column labels as A, T, G, C repeatedly. But this need not be the case. These combinations can be shuffled random. A DNA sequence can be of any length. This provides space for encrypting message of any length. Since only basic properties are used this can be programmed in any common programming languages.

E. Conclusion

Position weight matrix is not used in encryption procedures. It normally represents DNA sequences and used to study properties related to them. In a position weight matrix, sum of all the columns are same. In the matrix

constructed also care has been taken so that all the sum of all the columns is L. Even if the message is received by a hacker, they would try to treat it as a matrix related to DNA details and it is difficult for anyone to decrypt the message until the procedure is known. So this message can be send in public domain and it is difficult to differentiate between the original position weight matrix and the encrypted one. So the proposed method is safe for data transfer.

REFERENCES

[1] http://spin.atomicobject.com/2014/11/20/encryption-symmetric-asymmetric hashing.

[2] Dimple, Encryption using different techniques, A Review, International Journal in Multidisciplinary and Academic Research. 2013, vol. 2, no. 1, pp 1 – 9.

[3] PRANAB Grag, JASWINDER Singh Dilawari, A Review Paper on Cryptography and Significance of Key Length, International Journal of Computer Science and Communication Engineering IJCSCE Special issue on "Emerging Trends in Engineering" ICETIE. 2012, pp. 88 – 91.

[4] http://www.academia.edu/662784/Literature_Review_on_Cryptography_and_ Network_ Security

[5] MARAM, Balajee, RAO, K Laxmanan, KUMAR, Y Ramesh. Encryption and Decryption Algorithms using 3 – D Matrices, International Journal of Advanced Research in Computer Science and Software Engineering. 2013, vol 3, no. 4, pp 352– 356.

[6] M. Yamuna, S. Rave Rohit, Promod Mazumdar, Avani Gupta, Text Encryption Using Matrices, International Journal of Application or Innovation in Engineering and Management. 2013, vol. 2, no. 3, pp 265 – 268.

[7] HARINANDAN Tunga, A New Polybite Shuffling Encryption and Decryption Algorithm Based on N Dimensional Encryption Decryption Matrix, International Journal of Emerging Technology and Advanced Engineering. 2012, vol 2, no. 2, pp 143 – 149.

[8] NISARGA Chand, SUBHAJIT Bhattacharyya, A Novel Approach for Encryption of Text Messages Using Play – Fair Cipher 6 by 6 Matrix with Four Iteration Steps, International Journal of Engineering Science and Innovative Technology. 2014, vol 3, no. 1, pp. 478 – 484.

[9] https://en.wikipedia.org/wiki/Position_weight_matrix.